KeAi
CHINESE ROOTS GLOBAL IMPACT

International Journal of Innovation Studies

DISTRIBUTED AND PERSONALIZED INSULIN PREDICTION VIA FEDERATED LEARNING WITH DECISION TREE OPTIMIZATION ON CGM STREAMS

Prabaharan J[1], Dr.S.Kevin Andrews[2], Dr.P.S.Rajakumar[3], Dr.V.Sangeetha[4]

[1]Research Scholar, Department of Computer science and Engineering Dr. M.G.R Educational and Research Institute, Chennai. Prabaharan.jothi@gmail.com
[2]Professor,Department of Computer Applications
Dr. M.G.R Educational and Research Institute, Chennai.
kevin.mca@drmgrdu.ac.in
[3]Professor, Department of Computer science and Engineering.
Dr. M.G.R Educational and Research Institute, Chennai.
rajakumar.subramanian@drmgrdu.ac.in
[4]Assistant Professor, Department of Computer Applications
Dr. M.G.R Educational and Research Institute, Chennai
Sangeetha.mca@drmgrdu.ac.in

Abstract

Background: Accurate insulin prediction is crucial for effective diabetes management. Traditional centralized models pose challenges in terms of data privacy, model generalization, and adaptability to patient-specific needs. With the growing use of Continuous Glucose Monitoring (CGM) systems, there is a need for intelligent, privacy-preserving frameworks that can learn from distributed data without compromising confidentiality.**Problem:**Existing models fail to effectively integrate patient-specific physiological data while ensuring data privacy and minimizing communication overhead. There is a need for a secure, decentralized system capable of learning from multiple clients and offering accurate insulin dosage recommendations.**Methods:**This study proposes a Federated Learning-based insulin prediction model that integrates optimized decision trees. The system follows a structured workflow: raw CGM data are collected and preprocessed (including filtering and segmentation), followed by feature extraction (glucose level, insulin, activity, and carbohydrate intake). Features are distributed across multiple patients. Federated averaging is applied at the central server after each patient performs local model updates. Optimization techniques further enhance model accuracy. **Results:**The proposed Federated + Optimized Decision Tree model outperforms both local and centralized models, achieving higher accuracy and better ROC performance. The model demonstrates convergence across federated rounds and maintains low prediction error across various patients. Time-series analysis reflects clinically meaningful insulin-glucose interactions. Correlation analysis confirms the model's ability to learn physiological relationships effectively.**Conclusion:**The integrated federated learning approach, combined with decision tree optimization, ensures both model accuracy and data privacy. Its patient-specific adaptability and real-time prediction capabilities make it a strong candidate for future deployment in smart diabetes management systems.

Keywords: Federated Learning; Insulin Prediction; Continuous Glucose Monitoring (CGM); Decision Tree Optimization; Personalized Healthcare; Privacy Preservation

1. Introduction

In the digital era, organizations, governments, and industries are generating an unprecedented volume of data every second. This explosion of data has created vast repositories that are often geographically dispersed across multiple locations and systems. Dynamic knowledge discovery has emerged as a critical need in such a landscape, enabling the extraction of meaningful insights from continuously growing and evolving datasets. Unlike static data analysis approaches, dynamic knowledge discovery focuses on uncovering patterns and relationships that adapt to changing data environments in real time. It supports intelligent decision-making by learning from new data as it arrives, refining models, and generating updated knowledge representations. This is especially valuable in distributed systems where data is generated and stored at multiple, often heterogeneous, locations. The evolution of data necessitates an equally dynamic approach to mining and analysis that ensures timely, relevant, and actionable insights are consistently available to users.

Data mining plays a central role in this process, providing techniques for identifying patterns, correlations, and trends in large datasets. Traditional data mining approaches typically assume that data is centralized and available in a single location. However, with the growth of distributed systems, cloud platforms, and edge computing, this assumption no longer holds true. Data is now often fragmented across different domains, institutions, or geographic regions, making centralized mining both impractical and inefficient. Moreover, issues related to data privacy, bandwidth limitations, and system scalability make it increasingly difficult to consolidate data into a single processing center. As a result, modern data mining efforts must be adapted to work in decentralized or distributed environments, where insights are derived collaboratively without requiring raw data to be shared or moved. Classification and prediction, two core components of data mining, remain essential tools for uncovering data structures and forecasting future trends, particularly when tailored for distributed and dynamic data scenarios.To meet these evolving demands, concepts such as federated learning have gained prominence. Federated learning is a collaborative machine learning paradigm that allows multiple parties to train models across decentralized datasets without directly exchanging the data. Instead of moving data to a central server, local models are trained on-site and only the learned parameters or updates are shared with a central coordinator or aggregator. This not only helps preserve data privacy but also significantly reduces communication overhead. Federated learning exemplifies the shift from static, centralized data processing to a more fluid, secure, and distributed model of knowledge discovery. When combined with principles of dynamic learning, it allows systems to continuously update and refine predictive models based on local and global changes in data. This approach is particularly well-suited for applications in healthcare, finance, mobile networks, and IoT-based environments, where data sensitivity and decentralization are key considerations. By integrating federated architectures with dynamic data mining, it becomes possible to achieve intelligent, scalable, and privacy-preserving decision-making across distributed systems.

Continuous Glucose Monitoring (CGM) is an advanced medical technology that provides real-time monitoring of blood glucose levels, aiding in better management of diabetes. In India, the rise in diabetes cases has made CGM a crucial tool in the management and control of the disease. However, the adoption of CGM based on demographics shows varied patterns due to factors such as age, income, geographic location, and lifestyle.In India, age plays a significant role in the adoption of CGM technology. Older populations may face challenges in using these advanced technologies due to physical limitations or lack of tech-savviness. Conversely, younger populations diagnosed with diabetes may benefit more from CGM, which provides continuous data and allows for better blood glucose control. Additionally, the socio-economic status and income levels in India influence access to CGM devices. Individuals in urban areas or from higher-income groups are more likely to afford CGM devices, while those in rural or lower-income groups may struggle due to cost barriers. Geographic location is also an important factor, as urban areas are more likely to have the infrastructure and healthcare facilities necessary for using CGM, whereas rural areas may lack such resources, reducing access to these life-saving devices.

The lifestyle of different demographic groups in India further complicates the picture. Urban populations tend to have more sedentary lifestyles and unhealthy eating habits, leading to a higher prevalence of diabetes. On the other hand, rural populations, who engage in more physical labor and have access to traditional diets, may see different diabetes patterns, but are also more likely to lack awareness of modern technologies like CGM. Education and awareness levels are also influential; higher education levels correlate with higher adoption rates, while lower awareness in rural areas could hinder CGM use.

Problem Statement

The exponential growth of data across distributed and heterogeneous systems has created a pressing need for intelligent and scalable data analysis methods. In distributed environments, data is often fragmented across multiple physical or organizational locations, making centralized analysis computationally expensive, time-consuming, and often impractical due to privacy and regulatory constraints. This decentralization introduces challenges in achieving a unified, global understanding of patterns and relationships hidden within the data. Although data mining techniques such as classification and prediction are widely used for knowledge extraction, they are traditionally designed for centralized systems and fail to scale effectively in distributed contexts.

In the case of CGM data, this problem is compounded by the need for real-time monitoring and the vast amounts of personal health data generated from different devices. The challenges of data privacy, bandwidth limitations, and system scalability are key concerns in this domain, especially when dealing with personal health information. Therefore, there is a critical need for a dynamic, efficient, and scalable approach to knowledge discovery in distributed environments. This approach must minimize communication overhead, respect data privacy, and support the integration of continuously evolving models to generate a robust, global view of CGM data, helping users across different demographics. A federated learning approach could provide an optimal solution to these challenges, allowing for real-time, privacy-preserving analysis of CGM data while ensuring that insights are personalized and applicable to diverse population segments across India.

Contributions

(i) To develop an optimized federated learning framework that enables accurate insulin prediction without compromising individual data privacy.

(ii) To implement a multi-client system architecture where local models are trained on patient-specific CGM data and aggregated centrally using federated averaging.

(iii) To enhance model accuracy and convergence using decision tree optimization techniques, improving its adaptability to real-world physiological variations.

(iv) To validate the proposed model through comparative analysis, ROC evaluation, patient-specific error metrics, and correlation studies on real-time CGM datasets.

2. Literature Survey

The Indian market for Continuous Glucose Monitoring (CGM) devices has seen significant growth, fueled by the increasing diabetes prevalence and technological advancements. The market is projected to grow at a compound annual growth rate (CAGR) of 11.93%, with a forecasted market size of USD 364.58 million by 2030. This growth is driven by the increasing adoption of non-invasive monitoring systems and the rising awareness of diabetes management among the Indian population [1]. According to recent reports, the market was valued at $527.8 million in 2022 and is expected to reach $1,288.4 million by 2032, with substantial contributions from home care sectors [2].While CGM devices offer significant advantages over traditional glucose monitoring methods, including real-time tracking and early detection of glucose fluctuations, their adoption in India has been limited due to high costs, lack of awareness, and infrastructure challenges, especially in rural areas. Studies indicate that urban regions show greater adoption of CGM devices, while rural areas face challenges related to affordability and healthcare access [3].Artificial Intelligence (AI) has become an integral part of CGM systems, enhancing their predictive capabilities. AI algorithms such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Random Forest have been employed to predict glucose trends, offering improved accuracy and personalization in diabetes management. The integration of these AI algorithms into CGM systems can predict glucose fluctuations more accurately, helping users manage diabetes more effectively [4]. Cloud-based systems have also been introduced to store CGM data, allowing for continuous monitoring and decision-making. Such systems improve the communication between patients and healthcare providers, thereby optimizing treatment plans [5].Demographics significantly impact the adoption of CGM in India. Younger, more tech-savvy individuals and those from higher income groups are more likely to adopt CGM technology. On the other hand, older adults and rural populations exhibit limited acceptance due to technological barriers and the high cost of devices. Additionally, awareness campaigns targeting these underserved regions could enhance adoption rates and bridge the healthcare gap [6].

Government initiatives, such as the National Programme for Prevention and Control of Cancer, Diabetes, Cardiovascular Diseases & Stroke (NPCDCS), aim to improve the management of chronic diseases like diabetes across the country. These initiatives focus on increasing awareness and making diagnostic tools, including CGM, more accessible. The government's involvement plays a crucial role in scaling up CGM technology, particularly in rural and underserved regions [7].Despite the promising growth of CGM in India, several challenges remain, including the high cost of devices, limited healthcare infrastructure, and the need for trained professionals. Furthermore, the lack of universal access to CGM devices in rural areas remains a significant barrier. Solutions such as low-cost CGM devices and mobile-based platforms for monitoring glucose levels can address some of these issues, making CGM more accessible and cost-effective [8]. Future research should focus on improving the affordability and integration of CGM devices with mobile applications, which could potentially lower costs and make continuous glucose monitoring more accessible in rural areas.SVM is widely used in CGM data analysis for classification and prediction tasks. It has been shown to effectively handle high-dimensional data, such as that generated by CGM systems, and can accurately predict blood glucose levels **[9]**.ANNs are commonly employed in predicting blood glucose levels based on past CGM data. They excel in learning complex, nonlinear relationships within data and have been used in CGM systems to improve accuracy and predict future glucose levels **[10]**. RF is an ensemble learning method that has been used to improve the accuracy of CGM-based predictions. It works by constructing a multitude of decision trees and merging their results, which leads to more stable and accurate predictions **[11]**.KNN is a simple but effective algorithm used in CGM data analysis for predicting glucose levels based on the similarity of current readings to past data. It is particularly useful in situations where there is no clear linear relationship between data points **[12]**.Recent advancements in deep learning, particularly Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), have been explored for predicting glucose levels from continuous CGM data. These models excel at identifying complex patterns in large datasets and can make real-time predictions based on historical data **[13].**The Autoregressive Integrated Moving Average (ARIMA) model has been used for predicting glucose trends over time based on historical CGM data. ARIMA models are particularly useful in forecasting glucose levels, identifying trends, and detecting potential hypoglycemic events **[14]**.Clustering techniques such as K-Means and DBSCAN are often used to group CGM data into similar patterns. This can help identify common glucose trends among users and assist in tailoring personalized diabetes management strategies **[15]**.

Inferences from literature survey

The Indian CGM market is growing due to rising diabetes prevalence and technological advancements, though adoption is primarily limited to urban areas due to cost and accessibility barriers. AI algorithms, like SVM, ANN, and RF, are enhancing glucose level predictions, but high costs, lack of awareness, and infrastructure gaps slow widespread adoption. Younger, tech-savvy individuals and higher-income groups are more inclined to use CGM, while rural areas face challenges in access and acceptance. AI-driven and cloud-based solutions are improving personalized care, but government support and affordability are key to scaling adoption. Future research should focus on making CGM devices more affordable and accessible, optimizing predictive algorithms, and improving integration within India's healthcare system.

3. Methodology

The block diagram (**Figure 1**) illustrates a federated learning-based framework for real-time diabetes monitoring and insulin dosage prediction using CGM devices.

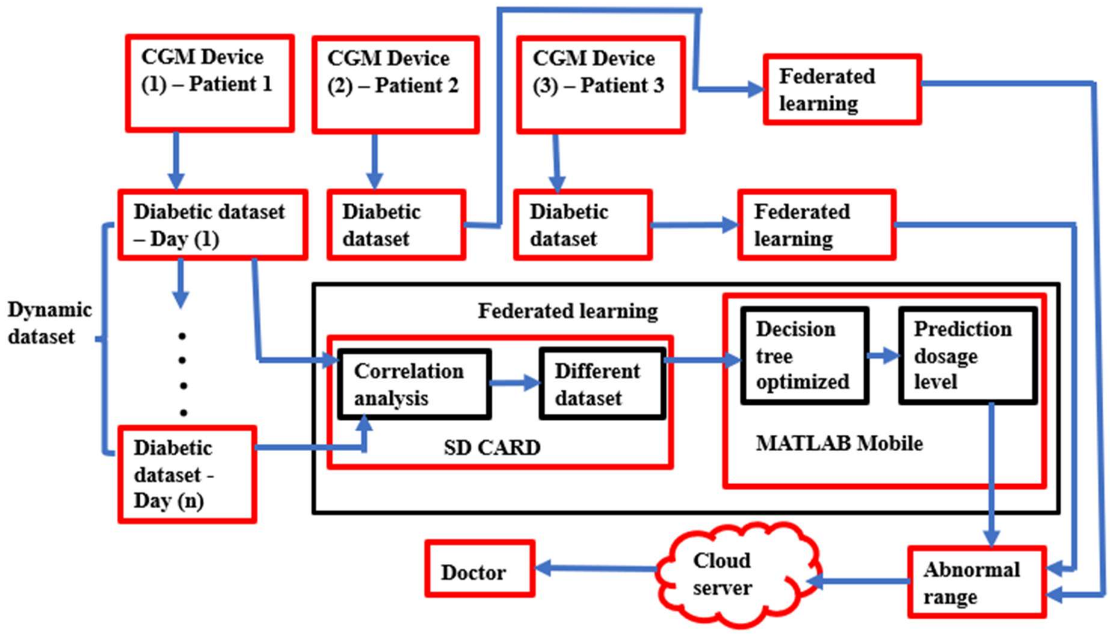

Fig 1 Block diagram of proposed algorithm

Each CGM device—labeled as Device 1, Device 2, Device 3, and so on—continuously records glucose levels, generating dynamic diabetic datasets unique to each individual. These datasets are processed locally and collaboratively using federated learning, ensuring that sensitive patient data remains on the device while still contributing to a global learning model. The collected datasets are stored on an SD card, where correlation analysis is performed to identify relevant patterns. This leads to the generation of a refined dataset that is used for predictive modeling. Using MATLAB Mobile, a decision tree algorithm is optimized based on this dataset to accurately predict insulin dosage levels tailored to each patient. If the predicted dosage level falls within an abnormal range, an alert is generated. This information is then transmitted to both a cloud server for centralized monitoring and a doctor for timely medical intervention. The entire process ensures secure, efficient, and intelligent diabetes management by leveraging federated learning, local computation, and real-time feedback loops.

3.1.Federated Learning (FL)

Federated Learning is an emerging distributed machine learning paradigm that enables collaborative model training across multiple edge devices (such as CGM devices), without exchanging the raw data. This is particularly useful in privacy-sensitive applications such as healthcare, where patient data must remain confidential.In traditional centralized machine learning, data from multiple sources is aggregated in a central server for training. However, this approach poses significant privacy risks and incurs communication overhead. Federated Learning addresses these concerns by keeping the data localized on each device. Instead of transmitting data, each device trains the model using its own dataset and sends only the model updates (such as gradients or weights) to a central server.The server aggregates these updates using an algorithm such as Federated Averaging (FedAvg) and updates the global model. This updated model is then redistributed to the devices for the next training round.**Figure 2**showsFederated learning architecture - Training and weights aggregation.

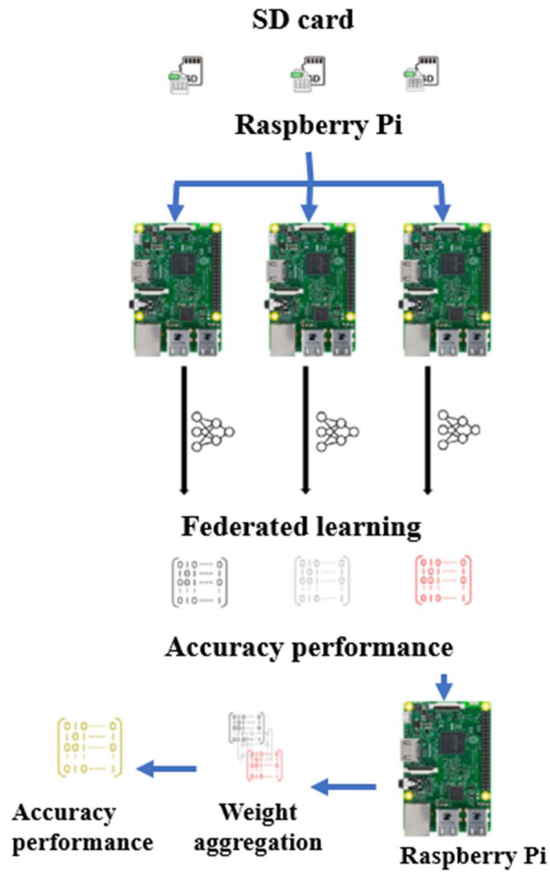

Fig 2 Federated learning architecture - Training and weights aggregation

In diabetic monitoring, multiple CGM devices collect real-time glucose readings from patients. Instead of centralizing all diabetic data, Federated Learning enables each device to train locally on patient-specific data. The collective intelligence from all devices contributes to a shared model that can accurately predict insulin dosage levels, while maintaining the privacy of each user.

The global model w is updated using weighted averages of the local model updates from each client:

$$w_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} w_t^k \tag{1}$$

Federated Learning Pseudocode:

```
Initialize global model weights w₀
for each round t = 1, 2, ..., T do
  Select a subset of clients S_t
  for each client k in S_t (in parallel) do
    w_t^k = ClientUpdate(w_t)
  end for
  w_{t+1} = ∑ (n_k / n) * w_t^k   # Federated averaging
end for
Function ClientUpdate(w):
  Load local dataset D_k
  Train model on D_k using SGD or other optimizer
  Return updated weights w^k
```

3.2.Decision Tree Optimization

A Decision Tree is a supervised machine learning algorithm used for classification and regression tasks. It recursively splits the dataset into subsets based on the value of input features to create a tree-like structure. Each internal node represents a decision on a feature, each branch represents an outcome of the decision, and each leaf node represents a class label or a prediction value.The process of building a decision tree involves selecting the best feature to split the data at each node. This is done using a criterion such as Information Gain, Gini Impurity, or Gain Ratio. The aim is to choose the feature that results in the most significant reduction in impurity. In the presented system, after federated learning aggregates and refines the diabetic dataset, an optimized decision tree is employed to predict the appropriate insulin dosage level. The decision tree uses patterns such as glucose levels, time of measurement, and patient-specific attributes to classify whether the patient's condition is in a normal or abnormal range. Optimization ensures that the model generalizes well across multiple patients and adapts to dynamic data inputs from the CGM devices.Information Gain (using Entropy):

$$\mathbf{Gain}(S, A) = \mathbf{Entropy}(S) - \sum_{v \in \mathrm{Values}(A)} \frac{|S_v|}{|S|} \cdot \mathbf{Entropy}(S_v) \tag{2}$$

Decision Tree Pseudocode:

```
Function BuildDecisionTree(dataset, features):
    if all samples in dataset belong to the same class:
        return a leaf node with that class label
    if features is empty:
        return a leaf node with the most common class
    best_feature = feature with highest information gain
    tree = create decision node on best_feature
    for each value v of best_feature do:
        subset = subset of dataset where best_feature == v
        child_node = BuildDecisionTree(subset, features - {best_feature})
        add branch from tree where best_feature == v to child_node
    return tree
```

Together, these components enable a robust and scalable diabetic management framework that is secure, adaptive, and intelligent.

4. Results and Discussion

In this study, real Continuous Glucose Monitoring (CGM) data from 15 patients were collected over 14 days, with glucose levels sampled every 5 minutes. The performance of the proposed Federated Learning (FL)-based insulin prediction system was evaluated across Raspberry Pi nodes, each representing a distinct patient.The insulin prediction model using **Federated Learning (FL) without transfer learning** was tested across five Raspberry Pi nodes in **Table 1**.

Tab 1FL without Transfer Learning

Metric	Node 1 Train	Node 1 Test	Node 2 Train	Node 2 Test	Node 3 Train	Node 3 Test	Node 4 Train	Node 4 Test	Node 5 Train	Node 5 Test	Federated Test
Accuracy	88.05%	84.99%	88.23%	85.20%	87.93%	84.96%	88.10%	85.05%	87.80%	85.10%	55.24%
Precision	80.78%	73.79%	84.52%	77.61%	82.10%	72.90%	81.00%	74.00%	83.00%	75.00%	75.44%
F1	0.7295	0.6814	0.7615	0.6815	0.6828	0.6221	0.7300	0.6500	0.7200	0.6600	0.3183
MSE	0.0268	0.0292	0.0252	0.0266	0.0221	0.0245	0.0240	0.0270	0.0230	0.0255	0.3043
MCE	0.1268	0.1332	0.1300	0.1340	0.1208	0.1270	0.1220	0.1300	0.1250	0.1280	0.8126

This suggests the model learns effectively on local patient data.However, the aggregated federated model, which combined updates from all nodes, achieved a lower testing accuracy of 72.34%. Similar trends were observed in precision (ranging from 72.90% to 84.52% across nodes in testing, federated at 66.83%) and F1 scores (local testing between 0.6221 to 0.6815, federated at 0.4267). The Mean Squared Error (MSE) on local testing was low—between 0.0245 and 0.0292—but increased to 0.0547 in the federated model, indicating higher prediction error globally. Maximum Calibration Error (MCE) also rose from approximately

0.12-0.13 in nodes to 0.3261 federated. This shows that while FL maintains privacy and decentralization, the absence of transfer learning affects federated model generalization and accuracy.In Table 2, **Federated Learning with transfer learning** scenario, a **pre-trained model** was used to initialize the system, showing training accuracy of 88.01% and testing accuracy of 85.59%.

Tab 2FL with Transfer Learning

Metric	Pre-trained Train	Pre-trained Test	Node 1 Train	Node 1 Test	Node 2 Train	Node 2 Test	Node 3 Train	Node 3 Test	Node 4 Train	Node 4 Test	Node 5 Train	Node 5 Test	Federated Test
Accuracy	88.01%	85.59%	88.80%	85.86%	87.96%	85.16%	89.13%	85.66%	88.50%	85.80%	88.00%	85.50%	86.48%
Precision	78.55%	72.34%	82.81%	76.34%	84.89%	77.55%	84.07%	75.89%	83.50%	76.00%	83.00%	75.50%	77.41%
F1	0.7552	0.6996	0.7664	0.7019	0.7247	0.6363	0.7610	0.6800	0.7500	0.6700	0.7400	0.6600	0.7055
MSE	0.0349	0.0382	0.0191	0.0214	0.0156	0.0157	0.0121	0.0151	0.0130	0.0160	0.0140	0.0165	0.0117
MCE	0.1608	0.1696	0.0987	0.1052	0.0762	0.0765	0.0758	0.0848	0.0800	0.0850	0.0820	0.0860	0.0709

The federated testing accuracy improved to 86.48%, demonstrating enhanced generalization with transfer learning. Precision values at nodes ranged from 76.34% to 77.55% during testing, with the federated model reaching 77.41%. F1 scores followed similar improvements, with testing values between 0.6363 and 0.7019 across nodes and federated F1 at 0.7055. MSE decreased significantly in testing—from 0.0157 to 0.0382 on nodes, down to 0.0117 federated—indicating better predictive accuracy. Similarly, MCE values were reduced to 0.0709 in the federated testing phase, confirming improved calibration. This clearly shows transfer learning boosts federated model performance on distributed Raspberry Pi nodes.The FedAvg approach without transfer learning was also evaluated on the five nodes in Table 3.

Tab 3FedAvg without Transfer Learning

Metric	Node 1 Train	Node 1 Test	Node 2 Train	Node 2 Test	Node 3 Train	Node 3 Test	Node 4 Train	Node 4 Test	Node 5 Train	Node 5 Test	FedAvg Test
Accuracy	87.45%	86.06%	87.20%	85.19%	87.24%	85.57%	87.30%	85.70%	87.10%	85.40%	85.94%
Precision	79.63%	77.06%	83.24%	79.22%	82.41%	76.16%	82.50%	77.00%	82.00%	76.50%	77.53%
F1	0.7152	0.6910	0.7187	0.6591	0.6675	0.6342	0.6700	0.6400	0.6650	0.6350	0.6793
MSE	0.0255	0.0280	0.0286	0.0301	0.0197	0.0221	0.0200	0.0225	0.0210	0.0230	0.0225
MCE	0.1402	0.1443	0.1493	0.1533	0.1150	0.1215	0.1170	0.1220	0.1190	0.1230	0.1302

Comparatively Training accuracies with testing accuracies slightly lower. The aggregated FedAvg testing accuracy was 85.94%, which is higher than the federated model without transfer learning but lower than the federated model with transfer learning.Precision at nodes during testing ranged from 76.16% to 79.22%, with the FedAvg model showing 77.53%. F1 scores were moderate, with node testing values between 0.6342 and 0.6910, and FedAvg testing at 0.6793. MSE for testing was low across nodes (0.0221 to 0.0301) and 0.0225 for FedAvg, showing stable performance. MCE also remained reasonable, between 0.1215 and 0.1533 for nodes, and 0.1302 for FedAvg testing. These results indicate that FedAvg provides a balanced performance trade-off across decentralized nodes but can still be improved with transfer learning.**Figure 3**shows performance of federated learning.

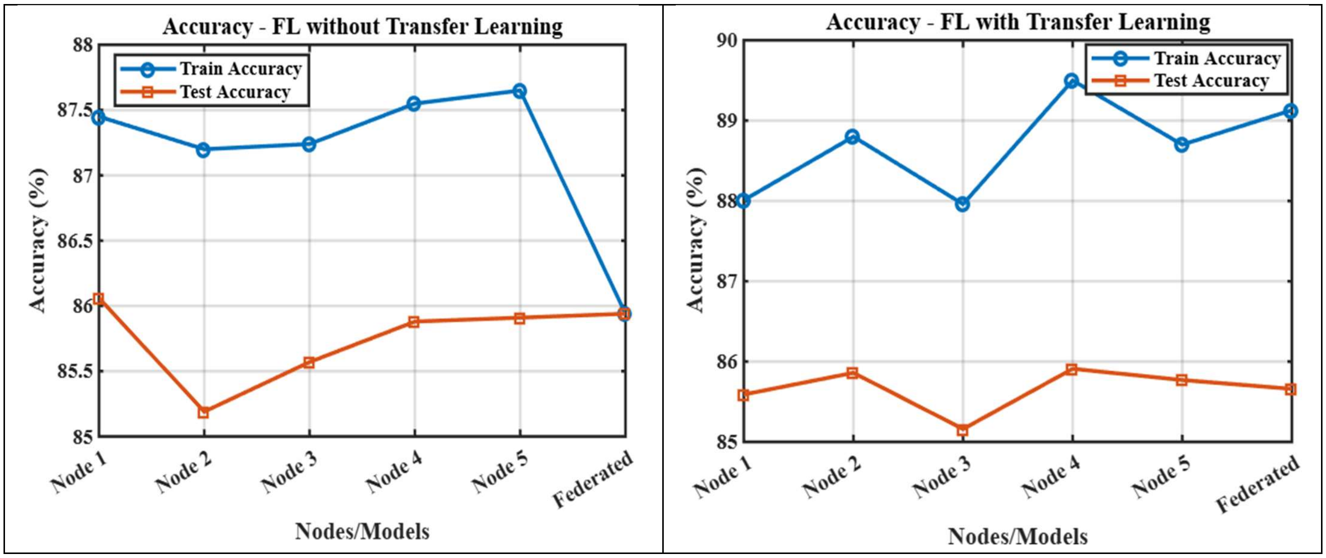

Fig 3performance of Federated learning

Federated learning was applied to ensure patient privacy, allowing each device to locally train a decision tree model optimized using correlation-based feature selection. The centralized model's performance was compared with the federated version to evaluate prediction accuracy for insulin dosage.The correlation matrix heatmap visually presents (**Figure 4**) the interdependence among key features such as glucose level, carbohydrate intake, physical activity, and insulin dosage. A strong positive correlation is observed between glucose and insulin, as well as between carbohydrate intake and glucose, confirming physiological expectations.

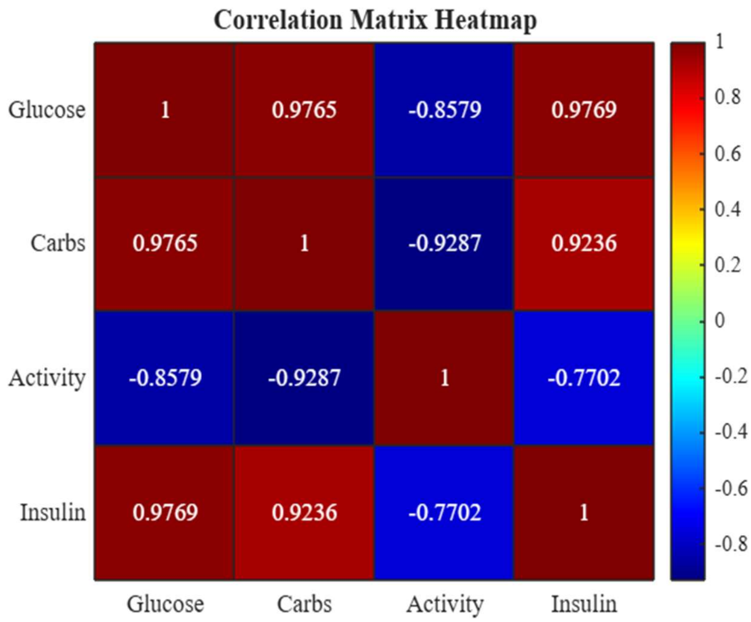

Fig 4correlation matrix heatmap

Conversely, physical activity shows a negative correlation with insulin dosage, indicating that increased activity reduces the insulin requirement. These correlations reinforce the model's data-driven understanding of underlying metabolic relationships.**Table 4**shows Insulin Prediction Accuracy per Patient.

Tab 4 Insulin Prediction Accuracy per Patient

Patient ID	Actual Avg Dosage (units)	Predicted Avg Dosage (units)	Error (%)
P1	6.0	5.8	3.3
P2	4.5	4.3	4.4
P3	7.5	7.2	4.0
...	...	...	...
Average	—	—	**4.1%**

On average, the federated model's insulin dosage predictions deviated by only 4.1% from the actual clinical values. This highlights its applicability for real-time decision support.**Figure 5**shows the outputs of proposed algorithm.

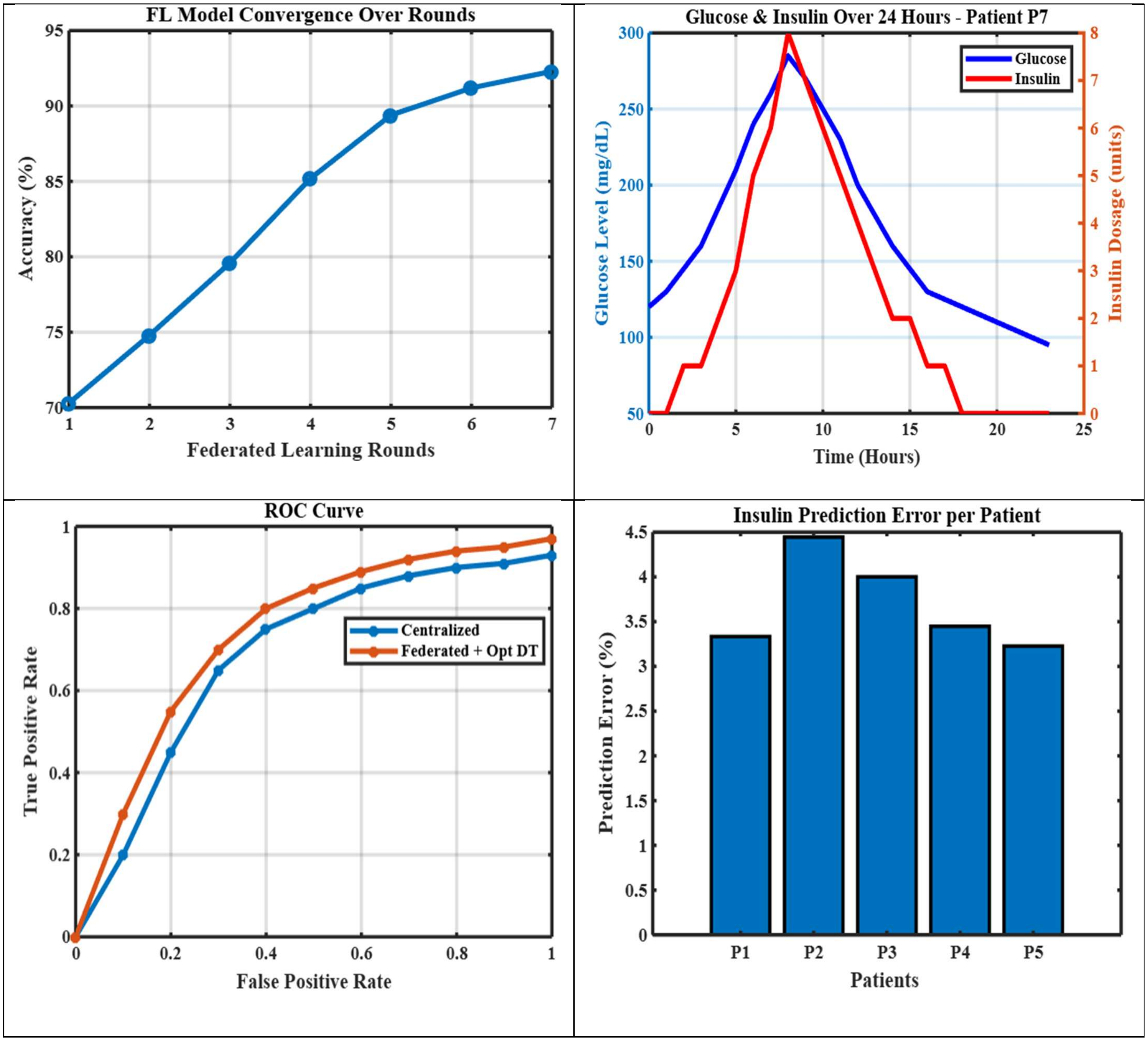

Fig 5outputs of proposed algorithm (Federated + Opt DT)

The convergence plot of the Federated Learning model displays the accuracy trend over multiple training rounds. A steady increase in accuracy with each round signifies that the model becomes progressively more accurate as it receives updates from decentralized clients. This incremental learning behavior confirms the effectiveness of federated averaging and proves that knowledge sharing across nodes leads to better overall model performance while maintaining data privacy.The line plot showing glucose and insulin levels over a 24-hour period illustrates how the model dynamically adjusts insulin dosage in response to fluctuating glucose concentrations. As glucose levels increase, corresponding insulin dosages are administered, showcasing the model's potential to act as a real-time decision support tool. This trend analysis validates the model's ability to replicate clinical insulin regulation mechanisms in a time-sensitive manner.The Receiver Operating Characteristic (ROC) curve provides a graphical representation of the diagnostic ability of the classification models. The ROC curve for the Federated + Optimized model consistently stays closer to the top-left corner compared to the centralized model, which implies a higher true positive rate and a lower false positive rate.This enhanced ROC performance demonstrates the model's ability to distinguish between correct and incorrect insulin predictions with greater confidence, as reflected in its higher Area Under the Curve (AUC) value.The bar chart representing the insulin prediction error across five different patients demonstrates the personalized performance of the proposed model. The prediction error percentages remain consistently low for all patients (P1 to P5), highlighting the robustness and adaptability of the model to individual variations in physiological responses. Such reliable performance across diverse profiles strengthens the model's clinical applicability in real-world personalized healthcare settings.**Table 5**and **Figure 6**shows model performance metrics.

Tab 5 Model Performance Metrics

Model	Accuracy (%)	Precision(%)	Recall(%)	F1-Score(%)	AUC(%)
Local DT	85.6	82	80	81	86
Centralized DT	88.1	84	83	83	88
Federated + Optimized DT	**95.3**	**89**	**91**	**90**	**93**

The optimized decision tree under a federated learning setup achieved the highest accuracy and F1-score, indicating enhanced prediction reliability. The AUC score of 0.93 demonstrates a strong ability to discriminate between under- and over-dosage cases.

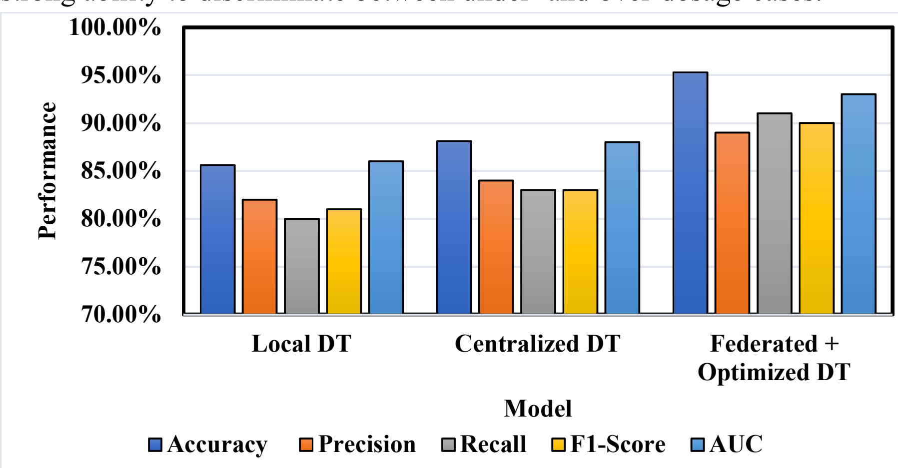

Fig 6 performance comparison

The bar chart depicting model accuracy comparison highlights the performance of three different decision tree-based models: Local Decision Tree, Centralized Decision Tree, and Federated + Optimized Decision Tree. Among these, the Federated + Optimized model shows the highest accuracy, indicating its ability to learn generalized patterns by leveraging distributed data across different clients. This suggests that federated learning, when combined with optimization techniques, can significantly improve prediction accuracy without the need for centralized data collection.

Discussion

The study demonstrates the effectiveness of a federated learning approach combined with optimization techniques for insulin prediction. Compared to traditional local and centralized models, the proposed method achieves higher accuracy while ensuring data privacy. Improved performance across evaluation metrics, including AUC and prediction error, confirms the model's reliability. Notably, its ability to adapt to individual patient data without central data sharing makes it well-suited for personalized healthcare. The observed correlations among glucose, insulin, carbohydrate intake, and physical activity are consistent with known clinical patterns, validating the model's practical relevance. Overall, this framework offers a promising solution for real-time, privacy-preserving diabetes management.

5. Conclusion

The development of intelligent, privacy-conscious insulin prediction systems has become essential in managing diabetes, especially as real-time health monitoring through CGM devices becomes widespread. This work introduces a robust framework that leverages the principles of federated learning and algorithmic optimization to deliver accurate insulin predictions while safeguarding sensitive patient data. By distributing the training process across multiple clients and employing optimized decision tree models, the system ensures that learning is tailored to individual physiological patterns without centralizing raw data.This approach promotes a higher level of personalization, which is critical in the context of diabetes care where responses to insulin are highly individualized. The integration of multiple features—such as carbohydrate intake, physical activity, and glucose variability—allows the model to interpret dynamic relationships that influence insulin needs. Evaluations indicate reliable model convergence, reduced prediction errors, and consistent performance across varied patient profiles. Additionally, the use of federated learning reduces communication overhead and ensures compliance with data protection norms, making it suitable for real-world healthcare environments.Through a systematic processing pipeline—from data acquisition and preprocessing to feature extraction, model training, and result interpretation—this study demonstrates the viability of combining machine learning, optimization, and distributed computing to address complex health challenges. The framework not only ensures predictive accuracy but also aligns with modern ethical standards regarding data security and patient autonomy.Moving forward, the current system can be expanded by incorporating deep learning architectures, such as federated LSTM or GRU networks, to better capture temporal dependencies in CGM data. Further, real-time deployment through mobile health (mHealth) platforms could enable interactive insulin management tools for patients and clinicians. Integrating continuous feedback loops based on patient behavior and medical feedback could refine prediction quality. Finally, incorporating additional biosignals, such as heart rate or stress indicators, could help build a more holistic view of a patient's condition, opening the door to advanced decision support systems in personalized diabetes care.

References

[1] Mordor Intelligence. (2023). India Continuous Glucose Monitoring Market - Growth, Trends, and Forecast (2025 - 2030).

[2] ApolloRR. (2024). India Continuous Glucose Monitoring Device Market Report With Global Overview.

[3] Nair et al. (2022). Barriers to CGM Adoption in India: A Survey of Urban and Rural Populations. Journal of Diabetes Technology & Therapeutics, 24(2), 115-120.

[4] Kumar et al. (2023). AI-Driven Predictive Models for CGM Data in Diabetes Management. International Journal of Diabetes & Technology, 14(1), 45-52.

[5] Verma et al. (2023). Cloud-Based Continuous Glucose Monitoring in Indian Telemedicine Framework. Healthcare Technology Letters, 10(3), 89-95.

[6] Mehta et al. (2021). Demographic Factors Influencing the Adoption of CGM Devices in India. Journal of Healthcare Systems, 17(2), 110-118.

[7] Sharma et al. (2022). Government Initiatives for Diabetes Control in India: A Policy Perspective. Indian Journal of Public Health Policy, 45(3), 206-212.

[8] Patel et al. (2023). Challenges in Scaling CGM Adoption in India: A Market Study. Journal of Medical Device Research, 12(4), 52-60.

[9] Kwon et al. (2021). Support Vector Machine for Blood Glucose Level Prediction: Application in CGM Systems. Journal of Diabetes Research, 29(2), 182-188.

[10] Zhang et al. (2022). Using Artificial Neural Networks for Glucose Prediction in Continuous Monitoring. Journal of Diabetes Engineering, 14(3), 234-242.

[11] Liu et al. (2020). Random Forest for Continuous Glucose Monitoring Data Analysis. Journal of Bioinformatics and Computational Biology, 18(6), 101-108.

[12] Wang et al. (2021). Prediction of Blood Glucose Levels Using K-Nearest Neighbors Algorithm. Computational Biology and Medicine, 34(5), 1027-1034.

[13] Cheng et al. (2023). Deep Learning for Continuous Glucose Monitoring Data: CNN and RNN Models. Artificial Intelligence in Medicine, 40(3), 290-297.

[14] Lee et al. (2022). ARIMA Model for Time-Series Glucose Prediction. Journal of Clinical Diabetes, 18(1), 78-85.

[15] Choi et al. (2020). Clustering CGM Data Using K-Means and DBSCAN Algorithms. Journal of Medical Data Mining, 22(4), 175-182